A-59941

# HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS

## FIELD OF THE INVENTION

The invention relates to *Haemophilus* adhesion and penetration proteins, nucleic acids, and vaccines.

## BACKGROUND OF THE INVENTION

Most bacterial diseases begin with colonization of a particular mucosal surface (Beachey et al., 1981, J. Infect. Dis. 143:325-345). Successful colonization requires that an organism overcome mechanical cleansing of the mucosal surface and evade the local immune response. The process of colonization is dependent upon specialized microbial factors that promote binding to host cells (Hultgren *et al.*, 1993 Cell, 73:887-901). In some cases the colonizing organism will subsequently enter (invade) these cells and survive intracellularly (Falkow, 1991, Cell 65:1099-1102).

*Haemophilus influenzae* is a common commensal organism of the human respiratory tract (Kuklinska and Kilian, 1984, Eur. J. Clin. Microbiol. 3:249-252). It is a human-specific organism that normally resides in the human nasopharynx and must colonize this site in order to avoid extinction. This microbe has a number of surface structures capable of promoting attachment to host cells (Guerina *et al.*, 1982, J. Infect. Dis. 146:564; Pichichero *et al.*, 1982, Lancet ii:960-962; St. Geme *et al.*, 1993, Proc. Natl. Acad. Sci. U.S.A.

90:2875-2879). In addition, *H. influenzae* has acquired the capacity to enter and survive within these cells (Forsgren *et al.*, 1994, Infect. Immun. 62:673-679; St. Geme and Falkow, 1990, Infect. Immun. 58:4036-4044; St. Geme and Falkow, 1991, Infect. Immun. 59:1325-1333, Infect. Immun. 59:3366-3371). As a result, this bacterium is an important cause of both localized respiratory tract and systemic disease (Turk, 1984, J. Med. Microbiol. 18:1-16). Nonencapsulated, non-typable strains account for the majority of local disease (Turk, 1984, supra); in contrast, serotype b strains, which express a capsule composed of a polymer of ribose and ribitol-5-phosphate (PRP), are responsible for over 95% of cases of *H. influenzae* systemic disease (Turk, 1982, Clinical importance of *Haemophilus influenzae*, p. 3-9. *In* S.H. Sell and P.F. Wright (ed.), *Haemophilus influenzae* epidemiology, immunology, and prevention of disease. Elsevier/North-Holland Publishing Co., New York).

The initial step in the pathogenesis of disease due to *H. influenzae* involves colonization of the upper respiratory mucosa (Murphy *et al.*, 1987, J. Infect. Dis. 5:723-731). Colonization with a particular strain may persist for weeks to months, and most individuals remain asymptomatic throughout this period (Spinola *et al.*, 1986, I. Infect. Dis. 154:100-109). However, in certain circumstances colonization will be followed by contiguous spread within the respiratory tract, resulting in local disease in the middle ear, the sinuses, the conjunctiva, or the lungs. Alternatively, on occasion bacteria will penetrate the nasopharyngeal epithelial barrier and enter the bloodstream.

*In vitro* observations and animal studies suggest that bacterial surface appendages called pili (or fimbriae) play an important role in *H. influenzae* colonization. In 1982 two groups reported a correlation between piliation and increased attachment to human oropharyngeal epithelial cells and erythrocytes (Guerina *et al.*, supra; Pichichero *et al.*, supra). Other investigators have demonstrated that anti-pilus antibodies block *in vitro* attachment by piliated *H. influenzae* (Forney *et al.*, 1992, J. Infect. Dis. 165:464-470; van Alphen *et al.*, 1988, Infect. Immun. 56:1800-1806). Recently Weber *et al.* insertionally inactivated the pilus structural gene in an *H. influenzae* type b strain and thereby eliminated expression of pili; the resulting mutant exhibited a reduced capacity for colonization of year-old monkeys (Weber *et al.*, 1991, Infect. Immun. 59:4724-4728).

A number of reports suggest that nonpilus factors also facilitate *Haemophilus* colonization. Using the human nasopharyngeal organ culture model, Farley *et al.* (1986, J. Infect. Dis. 161:274-280) and Loeb *et al.* (1988, Infect. Immun. 49:484-489) noted that nonpiliated type b strains were capable of mucosal attachment. Read and coworkers made similar observations upon examining nontypable strains in a model that employs nasal turbinate tissue in organ culture (1991, J. Infect. Dis. 163:549-558). In the monkey colonization study by Weber *et al.* (1991, supra), nonpiliated organisms retained a capacity for colonization, though at reduced densities; moreover, among monkeys originally infected with the piliated strain, virtually all organisms recovered from the nasopharynx were nonpiliated. All of these observations are consistent with the finding that nasopharyngeal isolates from children colonized with *H.*

*influenzae* are frequently nonpiliated (Mason *et al.*, 1985, Infect. Immun. 49:98-103; Brinton *et al.*, 1989, Pediatr. Infect. Dis. J. 8:554-561).

Previous studies have shown that *H. influenzae* are capable of entering (invading) cultured human epithelial cells via a pili-independent mechanism (St. Geme and Falkow, 1990, supra; St. Geme and Falkow, 1991, supra). Although *H. influenzae* is not generally considered an intracellular parasite, a recent report suggests that these *in vitro* findings may have an *in vivo* correlate (Forsgren *et al.*, 1994, supra). Forsgren and coworkers examined adenoids from 10 children who had their adenoids removed because of longstanding secretory otitis media or adenoidal hypertrophy. In all 10 cases there were viable intracellular *H. influenzae*. Electron microscopy demonstrated that these organisms were concentrated in the reticular crypt epithelium and in macrophage-like cells in the subepithelial layer of tissue. One possibility is that bacterial entry into host cells provides a mechanism for evasion of the local immune response, thereby allowing persistence in the respiratory tract.

Thus, a vaccine for the therapeutic and prophylactic treatment of *Haemophilus* infection is desirable. Accordingly, it is an object of the present invention to provide for recombinant Haemophilus Adherence and Penetration (HAP) proteins and variants thereof, and to produce useful quantities of these HAP proteins using recombinant DNA techniques.

It is a further object of the invention to provide recombinant nucleic acids encoding HAP proteins, and

expression vectors and host cells containing the nucleic acid encoding the HAP protein.

An additional object of the invention is to provide monoclonal antibodies for the diagnosis of *Haemophilus* infection.

A further object of the invention is to provide methods for producing the HAP proteins, and a vaccine comprising the HAP proteins of the present invention. Methods for the therapeutic and prophylactic treatment of *Haemophilus* infection are also provided.

## SUMMARY OF THE INVENTION

In accordance with the foregoing objects, the present invention provides recombinant HAP proteins, and isolated or recombinant nucleic acids which encode the HAP proteins of the present invention. Also provided are expression vectors which comprise DNA encoding a HAP protein operably linked to transcriptional and translational regulatory DNA, and host cells which contain the expression vectors.

The invention provides also provides methods for producing HAP proteins which comprises culturing a host cell transformed with an expression vector and causing expression of the nucleic acid encoding the HAP protein to produce a recombinant HAP protein.

The invention also includes vaccines for *Haemophilus influenzae* infection comprising an HAP protein for prophylactic or therapeutic use in generating an immune response in a patient. Methods of treating or

preventing *Haemophilus influenzae* infection comprise administering a vaccine.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B depict light micrographs of *H. influenzae* strains DB117(pGJB103) and DB117(pN187) incubated with Chang epithelial cells. Bacteria were incubated with an epithelial monolayer for 30 minutes before rinsing and straining with Giemsa stain. Figure 1A: *H. influenzae* strain DB117 carrying cloning vector alone (pGJB103); Figure 1B: *H. influenzae* strain DB117 harboring recombinant plasmid pH187. Bar represents 3.5 μm.

Figures 2A, 2B, 2C and 2D depict thin section transmission electron micrographs demonstrating interaction between *H. influenzae* strains N187 and DB117(pN187) with Chang epithelial cells. Bacteria were incubated with epithelial monolayers for four hours before rinsing and processing for examination by transmission electron microscopy. Figure 2A: strain N187 associated with the epithelial cell surface and present in an intracellular location; Figure 2B: *H. influenzae* DB117 (pH187) in intimate contact with the epithelial cell surface; Figure 2C: strain DB117(pN187) in the process of entering an epithelial cell; Figure 2D: strain DB117(pN187) present in an intracellular location. Bar represents 1 μm.

Figure 3 depicts outer membrane protein profiles of various strains. Outer membrane proteins were isolated on the basis of sarcosyl insolubility and resolved on a 10% SDS-polyacrylamide gel. Proteins were visualized by staining with Coomassie blue. Lane 1, *H. influenzae*

strain DB117(pGJB103); lane 2, strain DB117(pN187); lane 3, strain DB117(pJS106); lane 4, *E. coli* HB101(pGJB103); lane 5, HB101(pN187). Note novel proteins at ~160 kD and 45 kD marked by asterisks in lanes 2 and 3.

Figure 4 depicts a restriction map of pN187 and derivatives and locations of mini-Tn*10 kan* insertions. pN187 is a derivative of pGJB103 that contains an 8.5-kb *Sau*3AI fragment of chromosomal DNA from *H. influenzae* strain N187. Vector sequences are represented by hatched boxes. Letters above top horizontal line indicate restriction enzyme sites: Bg, *Bgl*II; C, *Cla*I; E, *Eco*RI; P, P*st*I. Numbers and lollipops above top horizontal line show positions of mini-Tn*10 kan* insertions; open lollipops represent insertions that have no effect on adherence and invasion, while closed lollipops indicate insertions that eliminate the capacity of pN187 to promote association with epithelial monolayers. Heavy horizontal line with arrow represents location of *hap* locus within pN187 and direction of transcription. (+): recombinant plasmids that promote adherence and invasion; (-): recombinant plasmids that fail to promote adherence and invasion.

Figure 5 depicts the identification of plasmid-encoded proteins using the bacteriophage T7 expression system. Bacteria were radiolabeled with [$^{35}$S] methionine, and whole cell lysates were resolved on a 10% SDS-polyacrylamide gel. Proteins were visualized by autoradiography. Lane 1, *E. coli* XL-1 Blue(pT7-7) uninduced; lane 2, XL-1 Blue(pT7-7) induced with IPTG; lane 3, XL-1 Blue(pJS103) uninduced; lane 4, XL-1 Blue(pJS103) induced with IPTG; lane 5, XL-1 Blue(pJS104) uninduced; lane 6, XL-1 Blue(pJS104) induced with IPTG. The plasmids pJS103 and pJS104 are

derivatives of pT7-7 that contain the 6.5-kb *Pst*I fragment from pN187 in opposite orientations. Asterisk indicates overexpressed protein in XL-1 Blue(pJS104).

Figures 6A, 6B, and 6C depict the nucleotide sequence and predicted amino acid sequence of hap gene. Putative -10 and -35 sequences 5′ to the *hap* coding sequence are underlined; a putative *rho*-independent terminator 3′ to the *hap* stop codon is indicated with inverted arrows. The first 25 amino acids of the protein, which are boxed, represent the signal sequence.

Figures 7A, 7B, 7C, 7D, 7E, 7F, 7G, and 7H depict a sequence comparison of the *hap* product and the cloned *H. influenzae* IgA1 proteases. Amino acid homologies between the deduced *hap* gene product and the *iga* gene products from *H. influenzae* HK368, HK61, HK393, and HK793 are shown. Dashes indicate gaps introduced in the sequences in order to obtain maximal homology. A consensus sequence for the five proteins is shown on the lower line. The conserved serine-type protease catalytic domain is underlined, and the common active site serine is denoted by an asterisk. The conserved cysteines are also indicated by asterisks.

Figure 8 depicts the IgA1 protease activity assay. Culture supernatants were assayed for the ability to cleave IgA1. Reaction mixtures were resolved on a 10% SDS-polyacrylamide gel and then transferred to a nitrocellulose membrane. The membrane was probed with antibody against human IgA1 heavy chain. Lane 1, *H. influenzae* strain N187; lane 2, strain DB117(pGJB103); lane 3, strain DB117(pN187). The cleavage product patterns suggest that strain N187 contains a type 2 IgA1 protease while strains DB117(pGJB103) and DB117(pN187)

contain a type 1 enzyme. The upper band of ~70-kD seen with the DB117 derivatives represents intact IgA1 heavy chain.

Figures 9A and 9B depict southern analysis of chromosomal DNA from strain *H. influenzae* N187, probing with *hap* versus *iga*. DNA fragments were separated on a 0.7% agarose gel and transferred bidirectionally to nitrocellulose membranes prior to probing with either *hap* or *iga*. Lane 1, N187 chromosomal DNA digested with *Eco*RI; lane 2, N187 chromosomal DNA digested with *Bgl*II; lane 3, N187 chromosomal DNA digested with *Bam*HI; lane 4, the 4.8-kb *Cla*I-*Pst*I fragment from pN187 that contains the intact *hap* gene. Figure 9A: Hybridization with the 4.8-kb *Cla*I-*Pst*I fragment containing the *hap* gene; Figure 9B: hybridization with the *iga* gene from *H. influenzae* strain Rd, carried as a 4.8-kb *Cla*I-*Eco*RI fragment in pVD116.

Figure 10 depicts a SDS-polyacrylamide gel of secreted proteins. Bacteria were grown to late log phase, and culture supernatants were precipitated with trichloroacetic acid and then resolved on a 10% SDS-polyacrylamide gel. Proteins were visualized by staining with Coomassie blue. Lane 1, *H. influenzae* strain DB117(pGJB103); lane 2, DB117(pN187); lane 3, DB117(pJS106); lane 4, DB117(pJS102); lane 5, DB117(pJS105); lane 6, DB117(Tn10-18); lane 7, DB117(Tn10-4'); lane 8, DB117(Tn10-30); lane 9, DB117(Tn10-16); lane 10, DB117(Tn10-10); lane Il, DB117(Tn10-8); lane 12, N187. Asterisk indicates 110-kD secreted protein encoded by *hap*.

## DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel *Haemophilus* Adhesion and Penetration (HAP) proteins. In a preferred embodiment, the HAP proteins are from *Haemophilus* strains, and in the preferred embodiment, from *Haemophilus influenza*. However, using the techniques outlined below, HAP proteins from other *Haemophilus influenzae* strains, or from other bacterial species such as *Neisseria* spp. or *Bordetalla* spp. may also be obtained.

A HAP protein may be identified in several ways. A HAP nucleic acid or HAP protein is initially identified by substantial nucleic acid and/or amino acid sequence homology to the sequences shown in Figure 6. Such homology can be based upon the overall nucleic acid or amino acid sequence.

The HAP proteins of the present invention have limited homology to *Haemophilus influenzae* and *N. gonorrhoeae* serine-type IgA1 proteases. This homology, shown in Figure 7, is approximately 30-35% at the amino acid level, with several stretches showing 55-60% identity, including amino acids 457-549, 399-466, 572-622, and 233-261. However, the homology between the HAP protein and the IgA1 protease is considerably lower than the similarity among the IgA1 proteases themselves.

In addition, the full length HAP protein has homology to Tsh, a hemagglutinin expressed by an avian *E. coli* strain (Provence and Curtiss 1994, Infect. Immun. 62:1369-1380). The homology is greatest in the N-terminal half of the proteins, and the overall homology is 30.5% homologous. The full length HAP protein also

has homology with pertactin, a 69 kD outer membrane protein expressed by *B. pertussis*, with the middle portion of the proteins showing 39% homology. Finally, HAP has 34 - 52% homology with six regions of HpmA, a calcium-independent hemolysin expressed by *Proteus mirabilis* (Uphoff and Welch, 1990, J. Bacteriol. 172:1206-1216).

As used herein, a protein is a "HAP protein" if the overall homology of the protein sequence to the amino acid sequence shown in Figure 6 is preferably greater than about 40 - 50%, more preferably greater than about 60% and most preferably greater than 80%. In some embodiments the homology will be as high as about 90 to 95 or 98%. This homology will be determined using standard techniques known in the art, such as the Best Fit sequence program described by Devereux *et al.*, *Nucl. Acid Res. 12*:387-395 (1984). The alignment may include the introduction of gaps in the sequences to be aligned. In addition, for sequences which contain either more or fewer amino acids than the protein shown in Figure 6, it is understood that the percentage of homology will be determined based on the number of homologous amino acids in relation to the total number of amino acids. Thus, for example, homology of sequences shorter than that shown in Figure 6, as discussed below, will be determined using the number of amino acids in the shorter sequence.

HAP proteins of the present invention may be shorter than the amino acid sequence shown in Figure 6. As shown in the Examples, the HAP protein may undergo post-translational processing similar to that seen for the serine-type IgA1 proteases expressed by *Haemophilus influenzae* and *N. gonorrhoeae*. These proteases are

synthesized as preproteins with three functional domains: the N-terminal signal peptide, the protease, and a C-terminal helper domain. Following movement of these proteins into the periplasmic space, the carboxy terminal ß-domain of the proenzyme is inserted into the outer membrane, possibly forming a pore (Poulsen *et al.*, 1989, Infect. Immun. 57:3097-3105; Pohlner *et al.*, 1987, Nature (London). 325:458-462; Klauser *et al.*, 1992, EMBO J. 11:2327-2335; Klauser *et al.*, 1993, J. Mol. Biol. 234:579-593). Subsequently the amino end of the protein is exported through the outer membrane, and autoproteolytic cleavage occurs to result in secretion of the mature 100 to 106-kD protease. The 45 to 56-kD C-terminal ß-domain remains associated with the outer membrane following the cleavage event. As shown in the Examples, the HAP nucleic acid is associated with expression of a 160 kD outer membrane protein. The secreted gene product is an approximately 110 kD protein, with the simultaneous appearance of a 45 kD outer membrane protein. The 45 kD protein appears to correspond to amino acids from about 960 to about 1394 of Figure 6. Any one of these proteins is considered a HAP protein for the purposes of this invention.

Thus, in a preferred embodiment, included within the defintion of HAP proteins are portions or fragments of the sequence shown in Figure 6. The fragments may be fragments of the entire sequence, the 110 kD sequence, or the 45 kD sequence. Generally, the HAP protein fragments may range in size from about 10 amino acids to about 1900 amino acids, with from about 50 to about 1000 amino acids being preferred, and from about 100 to about 500 amino acids also preferred. Particularly preferred fragments are sequences unique to HAP; these sequences have particular use in cloning HAP proteins

from other organisms or to generate antibodies specific to HAP proteins. Unique sequences are easily identified by those skilled in the art after examination of the HAP protein sequence and comparison to other proteins; for example, by examination of the sequence alignment shown in Figure 7. For instance, as compared to the IgA proteases, unique sequences include, but are not limited to, amino acids 11-14, 16-22, 108-120, 155-164, 257-265, 281-288, 318-336, 345-353, 398-416, 684-693, 712-718, 753-761, 871-913, 935-953, 985-1008, 1023-1034, 1067-1076, 1440-1048, 1585-1592, 1631-1639, 1637-1648, 1735-1743, 1863-1871, 1882-1891, 1929-1941, and 1958-1966 (using the numbering of Figure 7). HAP protein fragments which are included within the definition of a HAP protein include N- or C-terminal truncations and deletions which still allow the protein to be biologically active; for example, which still exhibit proteolytic activity in the case of the 110 kD putative protease sequence. In addition, when the HAP protein is to be used to generate antibodies, for example as a vaccine, the HAP protein must share at least one epitope or determinant with either the full length protein, the 110 kD protein or the 45 kD protein, shown in Figure 6. In a preferred embodiment, the epitope is unique to the HAP protein; that is, antibodies generated to a unique epitope exhibit little or no cross-reactivity with other proteins. By "epitope" or "determinant" herein is meant a portion of a protein which will generate and/or bind an antibody. Thus, in most instances, antibodies made to a smaller HAP protein will be able to bind to the full length protein.

In some embodiments, the fragment of the HAP protein used to generate antibodies are small; thus, they may

be used as haptens and coupled to protein carriers to generate antibodies, as is known in the art.

Preferably, the antibodies are generated to a portion of the HAP protein which remains attached to the *Haemophilus influenzae* organism. For example, the HAP protein can be used to vaccinate a patient to produce antibodies which upon exposure to the *Haemophilus influenzae* organism (e.g. during a subsequent infection) bind to the organism and allow an immune response. Thus, in one embodiment, the antibodies are generated to the roughly 45 kD fragment of the full length HAP protein. Preferably, the antibodies are generated to the portion of the 45 kD fragment which is exposed at the outer membrane.

In an alternative embodiment, the antibodies bind to the mature secreted 110 kD fragment. For example, as explained in detail below, the HAP proteins of the present invention may be administered therapeutically to generate neutralizing antibodies to the 110 kD putative protease, to decrease the undesirable effects of the 100 kD fragment.

In the case of the nucleic acid, the overall homology of the nucleic acid sequence is commensurate with amino acid homology but takes into account the degeneracy in the genetic code and codon bias of different organisms. Accordingly, the nucleic acid sequence homology may be either lower or higher than that of the protein sequence. Thus the homology of the nucleic acid sequence as compared to the nucleic acid sequence of Figure 6 is preferably greater than 40%, more preferably greater than about 60% and most preferably greater than

80%. In some embodiments the homology will be as high as about 90 to 95 or 98%.

In one embodiment, the nucleic acid homology is determined through hybridization studies. Thus, for example, nucleic acids which hybridize under high stringency to all or part of the nucleic acid sequence shown in Figure 6 are considered HAP protein genes. High stringency conditions include washes with 0.1XSSC at 65°C for 2 hours.

The HAP proteins and nucleic acids of the present invention are preferably recombinant. As used herein, "nucleic acid" may refer to either DNA or RNA, or molecules which contain both deoxy- and ribonucleotides. The nucleic acids include genomic DNA, cDNA and oligonucleotides including sense and anti-sense nucleic acids. Specifically included within the definition of nucleic acid are anti-sense nucleic acids. An anti-sense nucleic acid will hybridize to the corresponding non-coding strand of the nucleic acid sequence shown in Figure 6, but may contain ribonucleotides as well as deoxyribonucleotides. Generally, anti-sense nucleic acids function to prevent expression of mRNA, such that a HAP protein is not made, or made at reduced levels. The nucleic acid may be double stranded, single stranded, or contain portions of both double stranded or single stranded sequence. By the term "recombinant nucleic acid" herein is meant nucleic acid, originally formed <u>in vitro</u> by the manipulation of nucleic acid by endonucleases, in a form not normally found in nature. Thus an isolated HAP protein gene, in a linear form, or an expression vector formed <u>in vitro</u> by ligating DNA molecules that are not normally joined, are both considered recombinant for the purposes of this

invention. It is understood that once a recombinant nucleic acid is made and reintroduced into a host cell or organism, it will replicate non-recombinantly, i.e. using the in vivo cellular machinery of the host cell rather than in vitro manipulations; however, such nucleic acids, once produced recombinantly, although subsequently replicated non-recombinantly, are still considered recombinant for the purposes of the invention.

Similarly, a "recombinant protein" is a protein made using recombinant techniques, i.e. through the expression of a recombinant nucleic acid as depicted above. A recombinant protein is distinguished from naturally occurring protein by at least one or more characteristics. For example, the protein may be isolated away from some or all of the proteins and compounds with which it is normally associated in its wild type host, or found in the absence of the host cells themselves. Thus, the protein may be partially or substantially purified. The definition includes the production of a HAP protein from one organism in a different organism or host cell. Alternatively, the protein may be made at a significantly higher concentration than is normally seen, through the use of a inducible promoter or high expression promoter, such that the protein is made at increased concentration levels. Alternatively, the protein may be in a form not normally found in nature, as in the addition of an epitope tag or amino acid substitutions, insertions and deletions.

Also included with the definition of HAP protein are HAP proteins from other organisms, which are cloned and expressed as outlined below.

In the case of anti-sense nucleic acids, an anti-sense nucleic acid is defined as one which will hybridize to all or part of the corresponding non-coding sequence of the sequence shown in Figure 6. Generally, the hybridization conditions used for the determination of anti-sense hybridization will be high stringency conditions, such as 0.1XSSC at 65°C.

Once the HAP protein nucleic acid is identified, it can be cloned and, if necessary, its constituent parts recombined to form the entire HAP protein nucleic acid. Once isolated from its natural source, e.g., contained within a plasmid or other vector or excised therefrom as a linear nucleic acid segment, the recombinant HAP protein nucleic acid can be further used as a probe to identify and isolate other HAP protein nucleic acids. It can also be used as a "precursor" nucleic acid to make modified or variant HAP protein nucleic acids and proteins.

Using the nucleic acids of the present invention which encode HAP protein, a variety of expression vectors are made. The expression vectors may be either self-replicating extrachromosomal vectors or vectors which integrate into a host genome. Generally, these expression vectors include transcriptional and translational regulatory nucleic acid operably linked to the nucleic acid encoding the HAP protein. "Operably linked" in this context means that the transcriptional and translational regulatory DNA is positioned relative to the coding sequence of the HAP protein in such a manner that transcription is initiated. Generally, this will mean that the promoter and transcriptional initiation or start sequences are positioned 5′ to the HAP protein coding region. The transcriptional and

translational regulatory nucleic acid will generally be appropriate to the host cell used to express the HAP protein; for example, transcriptional and translational regulatory nucleic acid sequences from Bacillus will be used to express the HAP protein in Bacillus. Numerous types of appropriate expression vectors, and suitable regulatory sequences are known in the art for a variety of host cells.

In general, the transcriptional and translational regulatory sequences may include, but are not limited to, promoter sequences, leader or signal sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences. In a preferred embodiment, the regulatory sequences include a promoter and transcriptional start and stop sequences.

Promoter sequences encode either constitutive or inducible promoters. The promoters may be either naturally occurring promoters or hybrid promoters. Hybrid promoters, which combine elements of more than one promoter, are also known in the art, and are useful in the present invention.

In addition, the expression vector may comprise additional elements. For example, the expression vector may have two replication systems, thus allowing it to be maintained in two organisms, for example in mammalian or insect cells for expression and in a procaryotic host for cloning and amplification. Furthermore, for integrating expression vectors, the expression vector contains at least one sequence homologous to the host cell genome, and preferably two homologous sequences which flank the expression construct. The integrating

vector may be directed to a specific locus in the host cell by selecting the appropriate homologous sequence for inclusion in the vector. Constructs for integrating vectors are well known in the art.

In addition, in a preferred embodiment, the expression vector contains a selectable marker gene to allow the selection of transformed host cells. Selection genes are well known in the art and will vary with the host cell used.

The HAP proteins of the present invention are produced by culturing a host cell transformed with an expression vector containing nucleic acid encoding a HAP protein, under the appropriate conditions to induce or cause expression of the HAP protein. The conditions appropriate for HAP protein expression will vary with the choice of the expression vector and the host cell, and will be easily ascertained by one skilled in the art through routine experimentation. For example, the use of constitutive promoters in the expression vector will require optimizing the growth and proliferation of the host cell, while the use of an inducible promoter requires the appropriate growth conditions for induction. In addition, in some embodiments, the timing of the harvest is important. For example, the baculoviral systems used in insect cell expression are lytic viruses, and thus harvest time selection can be crucial for product yield.

Appropriate host cells include yeast, bacteria, archebacteria, fungi, and insect and animal cells, including mammalian cells. Of particular interest are Drosophila melangaster cells, Saccharomyces cerevisiae and other yeasts, E. coli, Bacillus subtilis, SF9 cells,

C129 cells, 293 cells, Neurospora, BHK, CHO, COS, and HeLa cells, immortalized mammalian myeloid and lymphoid cell lines.

In a preferred embodiment, HAP proteins are expressed in bacterial systems. Bacterial expression systems are well known in the art.

A suitable bacterial promoter is any nucleic acid sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of the coding sequence of HAP protein into mRNA. A bacterial promoter has a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region typically includes an RNA polymerase binding site and a transcription initiation site. Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose and maltose, and sequences derived from biosynthetic enzymes such as tryptophan. Promoters from bacteriophage may also be used and are known in the art. In addition, synthetic promoters and hybrid promoters are also useful; for example, the *tac* promoter is a hybrid of the *trp* and *lac* promoter sequences. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription.

In addition to a functioning promoter sequence, an efficient ribosome binding site is desirable. In *E. coli*, the ribosome binding site is called the Shine-Delgarno (SD) sequence and includes an initiation codon

and a sequence 3-9 nucleotides in length located 3 - 11 nucleotides upstream of the initiation codon.

The expression vector may also include a signal peptide sequence that provides for secretion of the HAP protein in bacteria. The signal sequence typically encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell, as is well known in the art. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria).

The bacterial expression vector may also include a selectable marker gene to allow for the selection of bacterial strains that have been transformed. Suitable selection genes include genes which render the bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin, neomycin and tetracycline. Selectable markers also include biosynthetic genes, such as those in the histidine, tryptophan and leucine biosynthetic pathways.

These components are assembled into expression vectors. Expression vectors for bacteria are well known in the art, and include vectors for *Bacillus subtilis, E. coli, Streptococcus cremoris*, and *Streptococcus lividans*, among others.

The bacterial expression vectors are transformed into bacterial host cells using techniques well known in the art, such as calcium chloride treatment, electroporation, and others.

In one embodiment, HAP proteins are produced in insect cells. Expression vectors for the transformation of insect cells, and in particular, baculovirus-based expression vectors, are well known in the art. Briefly, baculovirus is a very large DNA virus which produces its coat protein at very high levels. Due to the size of the baculoviral genome, exogenous genes must be placed in the viral genome by recombination. Accordingly, the components of the expression system include: a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the HAP protein; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene into the baculovirus genome); and appropriate insect host cells and growth media.

Mammalian expression systems are also known in the art and are used in one embodiment. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3′) transcription of a coding sequence for HAP protein into mRNA. A promoter will have a transcription initiating region, which is usually place proximal to the 5′ end of the coding sequence, and a TATA box, using a located 25-30 base pairs upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, typically located within 100 to 200 base pairs upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation. Of particular use as mammalian promoters

are the promoters from mammalian viral genes, since the viral genes are often highly expressed and have a broad host range. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter, and herpes simplex virus promoter.

Typically, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-translational cleavage and polyadenylation. Examples of transcription terminator and polyadenlytion signals include those derived form SV40.

The methods of introducing exogenous nucleic acid into mammalian hosts, as well as other hosts, is well known in the art, and will vary with the host cell used. Techniques include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

In a preferred embodiment, HAP protein is produced in yeast cells. Yeast expression systems are well known in the art, and include expression vectors for Saccharomyces cerevisiae, Candida albicans and C. maltosa, Hansenula polymorpha, Kluyveromyces fragilis and K. lactis, Pichia guillerimondii and P. pastoris, Schizosaccharomyces pombe, and Yarrowia lipolytica. Preferred promoter sequences for expression in yeast include the inducible GAL1,10 promoter, the promoters from alcohol dehydrogenase, enolase, glucokinase,

glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase, hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, pyruvate kinase, and the acid phosphatase gene. Yeast selectable markers include ADE2, HIS4, LEU2, TRP1, and ALG7, which confers resistance to tunicamycin; the G418 resistance gene, which confers resistance to G418; and the CUP1 gene, which allows yeast to grow in the presence of copper ions.

A recombinant HAP protein may be expressed intracellularly or secreted. The HAP protein may also be made as a fusion protein, using techniques well known in the art. Thus, for example, if the desired epitope is small, the HAP protein may be fused to a carrier protein to form an immunogen. Alternatively, the HAP protein may be made as a fusion protein to increase expression.

Also included within the definition of HAP proteins of the present invention are amino acid sequence variants. These variants fall into one or more of three classes: substitutional, insertional or deletional variants. These variants ordinarily are prepared by site specific mutagenesis of nucleotides in the DNA encoding the HAP protein, using cassette mutagenesis or other techniques well known in the art, to produce DNA encoding the variant, and thereafter expressing the DNA in recombinant cell culture as outlined above. However, variant HAP protein fragments having up to about 100-150 residues may be prepared by <u>in vitro</u> synthesis using established techniques. Amino acid sequence variants are characterized by the predetermined nature of the variation, a feature that sets them apart from naturally occurring allelic or interspecies variation of the HAP

protein amino acid sequence. The variants typically exhibit the same qualitative biological activity as the naturally occurring analogue, although variants can also be selected which have modified characteristics as will be more fully outlined below.

While the site or region for introducing an amino acid sequence variation is predetermined, the mutation per se need not be predetermined. For example, in order to optimize the performance of a mutation at a given site, random mutagenesis may be conducted at the target codon or region and the expressed HAP protein variants screened for the optimal combination of desired activity. Techniques for making substitution mutations at predetermined sites in DNA having a known sequence are well known, for example, M13 primer mutagenesis. Screening of the mutants is done using assays of HAP protein activities; for example, mutated HAP genes are placed in HAP deletion strains and tested for HAP activity, as disclosed herein. The creation of deletion strains, given a gene sequence, is known in the art. For example, nucleic acid encoding the variants may be expressed in a *Haemophilus influenzae* strain deficient in the HAP protein, and the adhesion and infectivity of the variant *Haemophilus influenzae* evaluated. Alternatively, the variant HAP protein may be expressed and its biological characteristics evaluated, for example its proteolytic activity.

Amino acid substitutions are typically of single residues; insertions usually will be on the order of from about 1 to 20 amino acids, although considerably larger insertions may be tolerated. Deletions range from about 1 to 30 residues, although in some cases

deletions may be much larger, as for example when one of the domains of the HAP protein is deleted.

Substitutions, deletions, insertions or any combination thereof may be used to arrive at a final derivative. Generally these changes are done on a few amino acids to minimize the alteration of the molecule. However, larger changes may be tolerated in certain circumstances.

When small alterations in the characteristics of the HAP protein are desired, substitutions are generally made in accordance with the following chart:

Chart I

| Original Residue | Exemplary Substitutions |
|---|---|
| Ala | Ser |
| Arg | Lys |
| Asn | Gln, His |
| Asp | Glu |
| Cys | Ser |
| Gln | Asn |
| Glu | Asp |
| Gly | Pro |
| His | Asn, Gln |
| Ile | Leu, Val |
| Leu | Ile, Val |
| Lys | Arg, Gln, Glu |
| Met | Leu, Ile |
| Phe | Met, Leu, Tyr |
| Ser | Thr |
| Thr | Ser |
| Trp | Tyr |
| Tyr | Trp, Phe |
| Val | Ile, Leu |

Substantial changes in function or immunological identity are made by selecting substitutions that are less conservative than those shown in Chart I. For

example, substitutions may be made which more significantly affect: the structure of the polypeptide backbone in the area of the alteration, for example the alpha-helical or beta-sheet structure; the charge or hydrophobicity of the molecule at the target site; or the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in the polypeptide's properties are those in which (a) a hydrophilic residue, e.g. seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g. leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g. lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g. glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g. phenylalanine, is substituted for (or by) one not having a side chain, e.g. glycine.

The variants typically exhibit the same qualitative biological activity and will elicit the same immune response as the naturally-occurring analogue, although variants also are selected to modify the characteristics of the polypeptide as needed. Alternatively, the variant may be designed such that the biological activity of the HAP protein is altered. For example, the proteolytic activity of the larger 110 kD domain of the HAP protein may be altered, through the substitution of the amino acids of the active site. The putative catalytic domain of this protein is GDSGSPMF, with the first serine corresponding to the active site serine characteristic of serine type proteases. The residues of the active site may be individually or simultaneously altered to decrease or eliminate proteolytic activity. This may be done to decrease the toxicity or side

effects of the vaccine. Similarly, the cleavage site between the 45 kD domain and the 100 kD domain may be altered, for example to eliminate proteolytic processing to form the two domains. Putatively this site is at residue 960.

In a preferred embodiment, the HAP protein is purified or isolated after expression. HAP proteins may be isolated or purified in a variety of ways known to those skilled in the art depending on what other components are present in the sample. Standard purification methods include electrophoretic, molecular, immunological and chromatographic techniques, including ion exchange, hydrophobic, affinity, and reverse-phase HPLC chromatography, and chromatofocusing. For example, the HAP protein may be purified using a standard anti-HAP antibody column. Ultrafiltration and diafiltration techniques, in conjunction with protein concentration, are also useful. For general guidance in suitable purification techniques, see Scopes, R., Protein Purification, Springer-Verlag, NY (1982). The degree of purification necessary will vary depending on the use of the HAP protein. In some instances no purification will be necessary.

Once expressed and purified if necessary, the HAP proteins are useful in a number of applications.

For example, the HAP proteins can be coupled, using standard technology, to affinity chromatography columns. These columns may then be used to purify antibodies from samples obtained from animals or patients exposed to the *Haemophilus influenzae* organism. The purified antibodies may then be used as outlined below.

Additionally, the HAP proteins are useful to make antibodies to HAP proteins. These antibodies find use in a number of applications. In a preferred embodiment, the antibodies are used to diagnose the presence of an *Haemophilus influenzae* infection in a sample or patient. This will be done using techniques well known in the art; for example, samples such as blood or tissue samples may be obtained from a patient and tested for reactivity with the antibodies, for example using standard techniques such as ELISA. In a preferred embodiment, monoclonal antibodies are generated to the HAP protein, using techniques well known in the art. As outlined above, the antibodies may be generated to the full length HAP protein, or a portion of the HAP protein.

Antibodies generated to HAP proteins may also be used in passive immunization treatments, as is known in the art.

Antibodies generated to unique sequences of HAP proteins may also be used to screen expression libraries from other organisms to find, and subsequently clone, HAP nucleic acids from other organisms.

In one embodiment, the antibodies may be directly or indirectly labelled. By "labelled" herein is meant a compound that has at least one element, isotope or chemical compound attached to enable the detection of the compound. In general, labels fall into three classes: a) isotopic labels, which may be radioactive or heavy isotopes; b) immune labels, which may be antibodies or antigens; and c) colored or fluorescent dyes. The labels may be incorporated into the compound at any position. Thus, for example, the HAP protein

antibody may be labelled for detection, or a secondary antibody to the HAP protein antibody may be created and labelled.

In one embodiment, the antibodies generated to the HAP proteins of the present invention are used to purify or separate HAP proteins or the *Haemophilus influenzae* organism from a sample. Thus for example, antibodies generated to HAP proteins which will bind to the *Haemophilus influenzae* organism may be coupled, using standard technology, to affinity chromatography columns. These columns can be used to pull out the *Haemophilus* organism from environmental or tissue samples. Alternatively, antibodies generated to the soluble 110 kD portion of the full-length portion of the protein shown in Figure 7 may be used to purify the 110 kD protein from samples.

In a preferred embodiment, the HAP proteins of the present invention are used as vaccines for the prophylactic or therapeutic treatment of a *Haemophilus influenzae* infection in a patient. By "vaccine" herein is meant an antigen or compound which elicits an immune response in an animal or patient. The vaccine may be administered prophylactically, for example to a patient never previously exposed to the antigen, such that subsequent infection by the *Haemophilus influenzae* organism is prevented. Alternatively, the vaccine may be administered therapeutically to a patient previously exposed or infected by the *Haemophilus influenzae* organism. While infection cannot be prevented, in this case an immune response is generated which allows the patient's immune system to more effectively combat the infection. Thus, for example, there may be a decrease or lessening of the symptoms associated with infection.

A "patient" for the purposes of the present invention includes both humans and other animals and organisms. Thus the methods are applicable to both human therapy and veterinary applications.

The administration of the HAP protein as a vaccine is done in a variety of ways. Generally, the HAP proteins can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby therapeutically effective amounts of the HAP protein are combined in admixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation are well known in the art. Such compositions will contain an effective amount of the HAP protein together with a suitable amount of vehicle in order to prepare pharmaceutically acceptable compositions for effective administration to the host. The composition may include salts, buffers, carrier proteins such as serum albumin, targeting molecules to localize the HAP protein at the appropriate site or tissue within the organism, and other molecules. The composition may include adjuvants as well.

In one embodiment, the vaccine is administered as a single dose; that is, one dose is adequate to induce a sufficient immune response to prophylactically or therapeutically treat a *Haemophilus influenzae* infection. In alternate embodiments, the vaccine is administered as several doses over a period of time, as a primary vaccination and "booster" vaccinations.

By "therapeutically effective amounts" herein is meant an amount of the HAP protein which is sufficient to induce an immune response. This amount may be different depending on whether prophylactic or therapeutic

treatment is desired. Generally, this ranges from about 0.001 mg to about 1 gm, with a preferred range of about 0.05 to about , and the preferred dose being _____. These amounts may be adjusted if adjuvants are used.

The following examples serve to more fully describe the manner of using the above-described invention, as well as to set forth the best modes contemplated for carrying out various aspects of the invention. It is understood that these examples in no way serve to limit the true scope of this invention, but rather are presented for illustrative purposes.

## EXAMPLES

### Example 1
### Cloning of the HAP protein

**Bacterial Strains, plasmids, and phage.** *H. influenzae* strain N187 is a clinical isolate that was originally cultivated from the middle ear fluid of a child with acute otitis media. This strain was classified as nontypable based on the absence of agglutination with typing antisera for *H. influenzae* types a-f (Burroughs Wellcome) and the failure to hybridize with pU038, a plasmid that contains the entire *cap b* locus (Kroll and Moxon, 1988, J. Bacteriol. 170:859-864).

*H. influenzae* strain DB117 is a *recl* mutant of Rd, a capsule-deficient serotype d strain that has been in the laboratory for over 40 years (Alexander and Leidy, 1951, J. Exp. Med. 83:345-359); DB117 was obtained from G. Barcak (University of Maryland, Baltimore, MD) (Sellow *et al.*, 1968). DB117 is deficient for *in vitro* adherence and invasion, as assayed below.

*H. influenzae* strain 12 is the nontypable strain from which the genes encoding the HMW1 and HMW2 proteins were cloned (Barenkamp and Leininger, 1992, Infect. Immun. 60:1302-1313); HMW1 and HMW2 are the prototypic members of a family of nontypable *Haemophilus* antigenically-related high-molecular-weight adhesive proteins (St. Geme *et al.*, 1993).

*E. coli* HB101, which is nonadherent and noninvasive, has been previously described (Sambrook *et al.*, 1989, Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. ). *E. coli* DH5α was obtained from Bethesda Research Laboratories. *E. coli* MC1061 was obtained from H. Kimsey (Tufts University, Boston, MA). *E. coli* XL-1 Blue and the plasmid pBluescript KS- were obtained from Stratagene. Plasmid pT7-7 and phage mGP1-2 were provided by S. Tabor (Harvard Medical School, Boston, MA) (Tabor and Richardson, 1985, Proc. Natl. Acad. Sci. USA. 82:1074-1078). The *E. coli-Haemophilus* shuttle vector pGJB103 (Tomb *et al.*, 1989, Rd. J. Bacteriol. 171:3796-3802) and phage λ1105 (Way *et al.*, 1984, Gene. 32:3 69-379) were provided by G. Barcak (University of Maryland, Baltimore, MD). Plasmid pVD116 harbors the IgA1 protease gene from *H. influenzae* strain Rd (Koomey and Falkow, 1984, Infect. Immun. 43:101-107) and was obtained from M. Koomey (University of Michigan, Ann Arbor, MI).

**Growth conditions.** *H. influenzae* strains were grown as described (Anderson *et al.*, 1972, J. Clin. Invest. 51:31-38). They were stored at -80°C in brain heart infusion broth with 25% glycerol. *E. coli* strains were grown on LB agar or in LB broth. They were stored at -80°C in LB broth with 50% glycerol.

For *H. influenzae*, tetracycline was used in a concentration of 5 μg/ml and kanamycin was used in a concentration of 25 μg/ml. For *E. coli*, antibiotics were used in the following concentrations: tetracycline, 12.5 μg/ml; kanamycin, 50 μg/ml; ampicillin, 100 μg/ml.

**Recombinant DNA methods.** DNA ligations, restriction endonuclease digestions, and gel electrophoresis were performed according to standard techniques (Sambrook *et al.*, 1989, supra). Plasmids were introduced into *E. coli* strains by either chemical transformation or electroporation, as described (Sambrook et al, 1989, supra; Dower *et al.*, 1988, Nucleic Acids Res. 16:617-6145). In *H. influenzae* transformation was performed using the MIV method of Herriott *et al.* (1970, J. Bacteriol. 101:517-524), and electroporation was carried out using the protocol developed for *E. coli* (Dower *et al.*, 1988, supra).

**Construction of genomic library from *H. influenzae* strain N187.** High-molecular-weight chromosomal DNA was prepared from 3 ml of an overnight broth culture of *H. influenzae* N187 as previously described (Mekalanos, 1983, Cell. 35:253-263). Following partial digestion with *Sau*3AI, 8 to 12 kb fragments were eluted into DEAE paper (Schleicher & Schuell, Keene, H.H.) and then ligated to *Bgl*II-digested calf intestine phosphatase-treated pGJB103. The ligation mixture was electroporated into *H. influenzae* DB117, and transformants were selected on media containing tetracycline.

**Transposon mutagenesis.**

Mutagenesis of plasmid DNA was performed using the mini-Tn*10 kan* element described by Way *et al.* (1984, supra). Initially, the appropriate plasmid was introduced into *E. coli* MC1061. The resulting strain was infected with λ1105, which carries the mini-Tn*10 kan* transposon. Transductants were grown overnight in the presence of kanamycin and an antibiotic to select for the plasmid, and plasmid DNA was isolated using the alkaline lysis method. In order to recover plasmids containing a transposon insertion, plasmid DNA was electroporated into *E. coli* DH5α, plating on media containing kanamycin and the appropriate second antibiotic.

In order to establish more precisely the region of pN187 involved in promoting interaction with host cells, initially this plasmid was subjected to restriction endonuclease analysis. Subsequently, several subclones were constructed in the vector pGJB103 and were reintroduced into *H. influenzae* strain DB117. The resulting strains were then examined for adherence and invasion. As summarized in Figure 4, subclones containing either a 3.9-kb *Pst*I-*Bgl*II fragment (pJS105) or the adjoining 4.2-kb *Bgl*II fragment (pJS102) failed to confer the capacity to associate with Chang cells. In contrast, a subclone containing an insert that included portions of both of these fragments (pJS106) did promote interaction with epithelial monolayers. Transposon mutagenesis performed on pH187 confirmed that the flanking portions of the insert in this plasmid were not required for the adherent/invasive phenotype. On the other hand, a transposon insertion located adjacent to the *Bgl*II site in pJS106 eliminated adherence and invasion. An insertion between the second *Eco*RI and *Pst*I sites in this plasmid had a similar effect (Figure 4).

**Examination of plasmid-encoded proteins.**

In order to examine plasmid encoded proteins, relevant DNA was ligated into the bacteriophage T7 expression vector pT7-7, and the resulting construct was transformed into *E. coli* XL-1 Blue. Plasmid pT7-7 contains the T7 phage $\phi$10 promoter and ribosomal binding site upstream of a multiple cloning site (Tabor and Richardson, 1985, supra). The T7 promoter was induced by infection with the recombinant M13 phage mGP1-2 and addition of isopropyl-$\beta$-D-thiogalactopyranoside (final concentration, 1 mM). Phage mGP1-2 contains the gene encoding T7 RNA polymerase, which activates the $\phi$10 promoter in pT7-7 (Tabor and Richardson, 1985, supra).

Like DB117(pN187), strain DB117 carrying pJS106 expressed new outer membrane proteins 160-kD and 45-kD in size (Figure 3, lane 3). In order to examine whether the 6.5-kb insert in pJS106 actually encodes these proteins, this fragment of DNA was ligated into the bacteriophage T7 expression vector pT7-7. The resulting plasmid containing the insert in the same orientation as in pN187 was designated pJS104, and the plasmid with the insert in the opposite orientation was designated pJS103. Both pJS104, and p7S103 were introduced into *E. coli* XL-1 Blue, producing XL-1 Blue(pJS104) and XL-1 Blue(pJS103), respectively. As a negative control, pT7-7 was also transformed into XL-1 Blue. The T7 promoter was induced in these three strains by infection with the recombinant M13 phage mGP1-2 and addition of isopropyl-$\beta$-D-thiogalactopyranoside (final concentration, 1 mM), and induced proteins were detected using [$^{35}$5] methionine. As shown in Figure 5, induction of XL-1 Blue(pJS104) resulted in expression of a 160-kD protein and several smaller proteins which presumably represent degradation products. In contrast, when XL-1

Blue(pJS103) and XL-1 Blue(pT7-7) were induced, there was no expression of these proteins. There was no 45-kD protein induced in any of the three strains. This experiment suggested that the 6.5-kb insert present in pJS106 contains the structural gene for the 160-kD outer membrane protein identified in DB117(pJS106). On the other hand, this analysis failed to establish the origin of the 45-kD membrane protein expressed by DB117(pJS106).

**Adherence and invasion assays.**

Adherence and invasion assays were performed with Chang epithelial cells [Wong-Kilbourne derivative, clone 1-5c-4 (human conjunctiva)], which were seeded into wells of 24-well tissue culture plates as previously described (St. Geme and Falkow, 1990). Adherence was measured after incubating bacteria with epithelial monolayers for 30 minutes as described (St. Geme *et al.*, 1993). Invasion assays were carried out according to our original protocol and involved incubating bacteria with epithelial cells for four hours followed by treatment with gentamicin for two hours (100 μg/ml) (St. Geme and Falkow, 1990).

**Nucleotide sequence determination and analysis.**

Nucleotide sequence was determined using a Sequenase kit and double stranded plasmid template. DNA fragments were subcloned into pBluescript KS$^-$ and sequenced along both strands by primer walking. DNA sequence analysis was performed using the Genetics Computer Group (GCG) software package from the University of Wisconsin (Devereux *et al.*, 1984). Sequence similarity searches were carried out using the BLAST program of the National Center for Biotechnology Information (Altschul *et al.*, 1990, J. Mol. Biol. 215:403-410). The DNA sequence

described here will be deposited in the EMBL/GenBank/DDBJ Nucleotide Sequence Data Libraries.

Based on the our subcloning results, we reasoned that the central *Bgl*II site in pH187 was positioned within an open reading frame. Examination of a series of mini-Tn*10 kan* mutants supported this conclusion (Figure 4). Consequently, we sequenced DHA on either side of this *Bgl*II site and identified a 4182 bp gene, which we have designated *hap* for *Haemophilus* adherence and penetration (Figure 6). This gene encodes a 1394 amino acid polypeptide, which we have called Hap, with a calculated molecular mass of 155.4-kD, in good agreement with the molecular mass of the larger of the two novel outer membrane proteins expressed by DB117(pN187) and the protein expressed after induction of XL-1 Blue/pJS104. The *hap* gene has a G+C content of 39.1%, similar to the published estimate of 38.7% for the whole genome (Kilian, 1976, J. Gen. Microbiol. 93:9-62). Putative -10 and -35 promoter sequences are present upstream of the initiation codon. A consensus ribosomal binding site is lacking. A sequence similar to a *rho*-independent transcription terminator is present beginning 39 nucleotides beyond the stop codon and contains interrupted inverted repeats with the potential for forming a hairpin structure containing a loop of three bases and a stem of eight bases. Similar to the situation with typical *E. coli* terminators, this structure is followed by a stretch rich in T residues. Analysis of the predicted amino acid sequence suggested the presence of a 25 amino acid signal peptide at the amino terminus. This region has characteristics typical of procaryotic signal peptides, with three positive H-terminal charges, a central hydrophobic region, and alanine residues at positions 23 and 25 (-3 and -1

relative to the putative cleavage site) (von Heijne, 1984, J. Mol. Biol. 173:243-251).

**Comparison of the deduced amino acid sequence of Hap with other proteins.** A protein sequence similarity search was performed with the predicted amino acid sequence using the BLAST network service of the National Center for Biotechnology Information (Altschul *et al.*, 1990, supra). This search revealed homology with the IgA1 proteases of *H. influenzae* and *Neisseria gonorrhoeae*. Alignment of the derived amino acid sequences for the *hap* gene product and the IgA1 proteases from four different *H. influenzae* strains revealed homology across the extent of the proteins (Figure 7), with several stretches showing 55-60% identity and 70-80% similarity. Similar levels of homology were noted between the *hap* product and the IgA1 protease from *N. gonorrhoeae* strain MS11. This homology includes the region identified as the catalytic site of the IgA1 proteases, which is comprised of the sequence GDSGSPLF, where 2 is the active site serine characteristic of serine proteases (Brenner, 1988, Nature (London). 334:528-530; Poulsen *et al.*, 1992, J. Bacteriol. 174:2913-2921). In the case of Hap, the corresponding sequence is GDSGSPMF. The *hap* product also contains two cysteines corresponding to the cysteines proposed to be important in forming the catalytic domain of the IgA proteases (Pohlner *et al.*, 1987, supra). Overall there is 30-35% identity and 51-55% similarity between the *hap* gene product and the *H. influenzae* and *N. gonorrhoeae* IgA proteases.

The deduced amino acid sequence encoded by *hap* was also found to contain significant homology to Tsh, a hemagglutinin expressed by an avian *E. coli* strain

(Provence and Curtiss, 1994, supra). This homology extends throughout both proteins but is greatest in the H-terminal half of each. Overall the two proteins are 30.5% identical and 51.6% similar. Tsh is also synthesized as a preprotein and is secreted as a smaller form; like the IgA1 proteases and perhaps Hap, a carboxy terminal peptide remains associated with the outer membrane (D. Provence, personal communication). While this protein is presumed to have proteolytic activity, its substrate has not yet been determined. Interestingly, Tsh was first identified on the basis of its capacity to promote agglutination of erythrocytes. Thus Hap and Tsh are possibly the first members of a novel class of adhesive proteins that are processed analogously to the IgA1 proteases.

Homology was also noted with pertactin, a 69-kD outer membrane protein expressed by *B. pertussis* (Charles *et al.*, 1989, Proc. Natl. Acad. Sci. USA. 86:3554-3558). The middle portions of these two molecules are 39% identical and nearly 60% similar. This protein contains the amino acid triplet arginine-glycine-aspartic acid (RGD) and has been shown to promote attachment to cultured mammalian cells via this sequence (Leininger *et al.*, 1991, Proc. Natl. Acad. Sci. USA. 88:345-349). Although *Bordetella* species are not generally considered intracellular parasites, work by Ewanowich and coworkers indicates that these respiratory pathogens are capable of *in vitro* entry into human epithelial cells (Ewanowich *et al.*, 1989, Infect. Immun. 57:2698-2704; Ewanowich *et al.*, 1989, Infect. Immun. 57:1240-1247). Recently Leininger *et al.* reported that preincubation of epithelial monolayers with an RGD-containing peptide derived from the pertactin sequence specifically inhibited *B. pertussis* entry (Leininger *et al.*, 1992,

Infect. Immun. 60:2380-2385). In addition, these investigators found that coating of *Staphylococcus aureus* with purified pertactin resulted in more efficient *S. aureus* entry; the RGD-containing peptide from pertactin inhibited this pertactin-enhanced entry by 75%. Although the *hap* product lacks an RGD motif, it is possible that Hap and pertactin serve similar biologic functions for *H. influenzae* and *Bordetella* species, respectively.

Additional analysis revealed significant homology (34 to 52% identity, 42 to 70% similarity) with six regions of HpmA, a calcium-independent hemolysin expressed by *Proteus mirabilis* (Uphoff and Welch, 1990, supra).

**The *hap* locus is distinct from the *H. influenzae* IgA1 protease gene.**

Given the degree of similarity between the *hap* gene product and *H. influenzae* IgA1 protease, we wondered whether we had isolated the IgA1 protease gene of strain N187. To examine this possibility, we performed IgA1 protease activity assays. Among *H. influenzae* strains, two enzymatically distinct types of IgA1 protease have been found (Mulks *et al.*, 1982, J. Infect. Dis. 146:266-274). Type 1 enzymes cleave the Pro-Ser peptide bond between residues 231 and 232 in the hinge region of human IgA1 heavy chain and generate fragments of roughly 28-kD and 31-kD; type 2 enzymes cleave the Pro-Thr bond between residues 235 and 236 in the hinge region and generate 26.5-kD and 32.5-kD fragments. Previous studies of the parent strain from which DB117 was derived have demonstrated that this strain produces a type 1 IgA1 protease (Koomey and Falkow, 1984, supra). As shown in Figure 8, comparison of the proteolytic activities of strain DB117 and strain N187 suggested

that N187 produces a type 2 IgA1 protease. We reasoned that DB117(pN187) might generate a total of four fragments from IgA1 protease, consistent with two distinct cleavage specificities. Examination of DB117(pH187) revealed instead that this transformant produces the same two fragments of the IgA1 heavy chain as does DB117, arguing that this strain produces only a type 1 enzyme.

In an effort to obtain additional evidence against the possibility that plasmid pH187 contains the N187 IgA1 protease gene, we performed a series of Southern blots. As shown in Figure 9, when genomic DNA from strain N187 was digested with *Eco*RI, *Bgl*II, or *Bam*HI and then probed with the *hap* gene, one set of hybridizing fragments was detected. Probing of the same DNA with the *iga* gene from *H. influenzae* strain Rd resulted in a different set of hybridizing bands. Moreover, the *iga* gene failed to hybridize with a purified 4.8-kb fragment that contained the intact *hap* gene.

**The recombinant plasmid associated with adherence and invasion encodes a secreted protein.**

The striking homology between the *hap* gene product and the *Haemophilus* and *Neisseria* IgA1 proteases suggested the possibility that these proteins might be processed in a similar manner. The IgA1 proteases are synthesized as preproteins with three functional domains: the N-terminal signal peptide, the protease, and a C-terminal helper domain, which is postulated to form a pore in the outer membrane for secretion of the protease (Poulsen *et al.*, 1989, supra; Pohlner *et al.*, 1987, supra). The C-terminal peptide remains associated with the outer membrane following an autoproteolytic cleavage event that results in release of the mature enzyme.

Consistent with the possibility that the *hap* gene product follows a similar fate, we found that DB117(pN187) produced a secreted protein approximately 110-kD in size that was absent from DB117(pGJB103) (Figure 10). This protein was also produced by DB117(pJS106), but not by DB117(pJ5102) or DB117(pJS105). Furthermore, the two mutants with transposon insertions within the *hap* coding region were deficient in this protein. In order to determine the relationship between *hap* and the secreted protein, this protein was transferred to a PVDF membrane and N-terminal amino acid sequencing was performed. Excessive background on the first cycle precluded identification of the first amino acid residue of the free amino terminus. The sequence of the subsequent seven residues was found to be HTYFGID, which corresponds to amino acids 27 through 33 of the *hap* product.

The introduction of *hap* into laboratory strains of *E. coli* strains was unable to endow these organisms with the capacity for adherence or invasion. In considering these results, it is noteworthy that the *E. coli* transformants failed to express either the 160-kD or the 45-kD outer membrane protein. Accordingly, they also failed to express the 110-kD secreted protein. The explanation for this lack of expression is unclear. One possibility is that the *H. influenzae* promoter or ribosomal binding site was poorly recognized in *E. coli*. Indeed the putative -35 sequence upstream of the *hap* initiation codon is fairly divergent from the $\sigma 70$ consensus sequence, and the ribosomal binding site is unrecognizable. Alternatively, an accessory gene may be required for proper export of the Hap protein, although the striking homology with the IgA proteases,

which are normally expressed and secreted in *E. coli*, argues against this hypothesis.

In considering the possibility that the *hap* gene product promotes adherence and invasion by directly binding to a host cell surface structure, it seems curious that the mature protein is secreted from the organism. However, there are examples of other adherence factors that are also secreted. Filamentous hemagglutinin is a 220-kD protein expressed by *B. pertussis* that mediates *in vitro* adherence and facilitates natural colonization (Relman *et al.*, 1989, Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641; Kimura *et al.*, 1990, Infect. Immun. 58:7-16). This protein remains surface-associated to some extent but is also released from the cell. The process of Filamentous hemagglutinin secretion involves an accessory protein designated FhaC, which appears to be localized to the outer membrane (Willems *et al.*, 1994, Molec. Microbiol. 11:337-347). Similarly, the Ipa proteins implicated in Shigella invasion are also secreted. Secretion of these proteins requires the products of multiple genes within the *mxi* and *spa* loci (Allaoui *et al.*, 1993, Molec. Microbiol. 7:59-68; Andrews *et al.*, 1991, Infect. Immun. 59:1997-2005; Venkatsan *et al.*, 1992, J. Bacteriol. 174:1990-2001).

It is conceivable that secretion is simply a consequence of the mechanism for export of the *hap* gene product to the surface of the organism. However, it is noteworthy that the secreted protein contains a serine-type protease catalytic domain and shows homology with the *P. mirobilis* hemolysin. These findings suggest that the mature Hap protein may possess proteolytic activity and raise the possibility that Hap promotes interaction with the host cell at a distance by modifying the host cell

surface. Alternatively, Hap may modify the bacterial surface in order to facilitate interaction with a host cell receptor. It is possible that *hap* encodes a molecule with dual functions, serving as both adhesin and protease.

**Analysis of outer membrane and secreted proteins.** Outer membrane proteins were isolated on the basis of sarcosyl insolubility according to the method of Carlone *et al.* (1986, J. Clin. Microbiol. 24:330-332). Secreted proteins were isolated by centrifuging bacterial cultures at 16,000 g for 10 minutes, recovering the supernatant, and precipitating with trichloroacetic acid in a final concentration of 10%. SDS-polyacrylamide gel electrophoresis was performed as previously described (Laemmli, 1970, Nature (London). 227:680-685).

To identify proteins that might be involved in the interaction with the host cell surface, outer membrane protein profiles for DB117(pN187) and DB117(pGJB103) were compared. As shown in Figure 3, DB117(pN187) expressed two new outer membrane proteins: a high-molecular-weight protein approximately 160-kD in size and a 45-kD protein. *E. coli* HB101 harboring pN187 failed to express these proteins, suggesting an explanation for the observation that HB101(pN187) is incapable of adherence or invasion.

Previous studies have demonstrated that a family of antigenically-related high-molecular-weight proteins with similarity to filamentous hemagglutinin of *Bordetella pertussis* mediate attachment by nontypable *H. influenzae* to cultured epithelial cells (St. Geme *et al.*, 1993). To explore the possibility that the gene encoding the strain H187 member of this family was

cloned, whole cell lysates of N187, DB117(pN187), and DB117(pGJB103) were examined by Western immunoblot. Our control strain for this experiment was *H. influenzae* strain 12. Using a polyclonal antiserum directed against HMW1 and HMW2, the prototypic proteins in this family, we identified a 140-kD protein in strain H187 (not shown). In contrast, this antiserum failed to react with either DB117(pN187) or DB117(pGJB103) (not shown), indicating that pN187 has no relationship to HMW protein expression.

**Determination of amino terminal sequence.** Secreted proteins were precipitated with trichloroacetic acid, separated on a 10% SDS-polyacrylamide gel, and electrotransferred to a polyvinylidene difluoride (PVDF) membrane (Matsudaira, 1987, J. Biol. Chem. 262:10035-10038). Following staining with Coomassie Brilliant Blue R-250, the 110-kD protein was cut from the PVDF membrane and submitted to the Protein Chemistry Laboratory at Washington University School of Medicine for amino terminal sequence determination. Sequence analysis was performed by automated Edman degradation using an Applied Biosystems Model 470A protein sequencer.

**Examination of IgA1 protease activity.** In order to assess IgA1 protease activity, bacteria were inoculated into broth and grown aerobically overnight. Samples were then centrifuged in a microphage for two minutes, and supernatants were collected. A 10 $\mu$l volume of supernatant was mixed with 16 $\mu$l of 0.5 $\mu$g/ml human IgA1 (Calbiochem), and chloramphenicol was added to a final concentration of 2 $\mu$g/ml. After overnight incubation at 37°C, reaction mixtures were electrophoresed on a 10% SDS-polyacrylamide gel, transferred to a nitrocellulose

membrane, and probed with goat anti-human IgA1 heavy chain conjugated to alkaline phosphatase (Kirkegaard & Perry). The membrane was developed by immersion in phosphatase substrate solution (5-bromo-4-chloro-3-indolylphosphate toluidinium-nitro blue tetrazolium substrate system; Kirkegaard & Perry).

**Immunoblot analysis.** Immunoblot analysis of bacterial whole cell lysates was carried out as described (St. Geme *et al.*, 1991).

**Southern hybridization.** Southern blotting was performed using high stringency conditions as previously described (St. Geme and Falkow, 1991).

**Microscopy.**

*i. Light microscopy.* Samples of epithelial cells with associated bacteria were stained with Giemsa stain and examined by light microscopy as described (St. Geme and Falkow, 1990).

*ii. Transmission electron microscopy.* For transmission electron microscopy, bacteria were incubated with epithelial cell monolayers for four hours and were then rinsed four times with PBS, fixed with 2% glutaraldehyde/1% osmium tetroxide in 0.1 M sodium phosphate buffer pH 6.4 for two hours on ice, and stained with 0.25% aqueous uranyl acetate overnight. Samples were then dehydrated in graded ethanol solutions and embedded in polybed. Ultrathin sections (0.4 $\mu$m) were examined in a Phillips 201c electron microscope.

As shown in Figure 2, DB117(pN187) incubated with monolayers for four hours demonstrated intimate interaction with the epithelial cell surface and was

occasionally found to be intracellular. In a given thin section, invaded cells generally contained one or two intracellular organisms. Of note, intracellular bacteria were more common in sections prepared with strain N187, an observation consistent with results using the gentamicin assay. In contrast, examination of samples prepared with strain DB117 carrying cloning vector alone (pGJB103) failed to reveal internalized bacteria (not shown).

Having described the preferred embodiments of the present invention it will appear to those of ordinary skill in the art that various modifications may be made to the disclosed embodiments, and that such modifications are intended to be within the scope of the present invention.